DOCUMENT RESUME

ED 251 744 CG 017 872

AUTHOR Dossett, Dennis L.; Gier, Joseph A.
TITLE Employee Acceptance of BOS and BES Performance Appraisals.
PUB DATE Aug 84
NOTE 15p.; Paper presented at the Annual Convention of the American Psychological Association (92nd, Toronto, Ontario, Canada, August 24-28, 1984).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Adoption (Ideas); *Behavior Rating Scales; *Employee Attitudes; Evaluation Methods; Personnel Evaluation; Sales Workers
IDENTIFIERS Behavioral Expectation Scales; Behavioral Observation Scales; *Performance Appraisal

ABSTRACT

Previous research on performance evaluation systems has failed to take into account user acceptance. Employee acceptance of a behaviorally-based performance appraisal system was assessed in a field experiment contrasting user preference for Behavioral Expectations Scales (BES) versus Behavioral Observation Scales (BOS). Non-union sales associates (N=240) of a retail chain were rated by either the BOS or BES appraisal form. Following the performance rating, managers conducted a formal performance review with each sales associate, focusing on reinforcement for engaging in effective job behaviors and setting specific behavioral goals. Sixty-nine percent of the employees (N=165) then completed a questionnaire about the rating form and the performance review session. No significant differences were found between format conditions for self-ratings of goal specificity, participation in goal setting, or perceived goal difficulty. However significant differences in perceptions of fairness and utility were found as well as evidence suggesting greater self-perceptions of motivation and goal acceptance when appraisals were conducted with the BOS rating format. (JAC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED251744

Employee Acceptance of BOS and BES Performance Appraisals

Dennis L. Dossett
University of Missouri - St. Louis

Joseph A. Gier
University of Nebraska at Omaha

Running head: Employee Acceptance

Presented at the meeting of the American Psychological Association, Toronto, Canada, August 1984.

Joseph A. Gier is now with Zales Corporation, Dallas, Texas.

CG 017872

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Dennis L. Dossett

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Abstract

Employee acceptance of a behaviorally-based performance appraisal system was assessed in a field experiment contrasting user preference for behavioral expectation scales (BES) versus behavioral observation scales (BOS). It was hypothesized that the greater behavioral specificity of BOS would be perceived by employees as more accurate and fair, as having greater utility for coaching and development, and as being more motivating than BES. No significant differences between format conditions were found for the context factors of goal specificity, participation in goal setting, or perceived goal difficulty. Significant differences in perceptions of fairness and utility were found as well as evidence suggesting greater self-perceptions of motivation and goal acceptance when appraisals were conducted with the BOS rating format.

User Acceptance of BES and BOS Performance Appraisals

Jacobs, Kafry, and Zedeck (1980) outlined 11 utilization and quality criteria they considered to be requisite properties for a performance evaluation system. However, their excellent analysis is deficient with respect to one important dimension, user acceptance. Lawler (1967) recognized this factor, including a variable which he called "attitudes toward fairness and acceptability of the system", in a model explicating the variables which affect the validity of performance rating systems. An otherwise perfect evaluation system will fail if supervisors and employees do not perceive it to be accurate, fair, and supportive of the best interests of both the organization and the employees. Supervisors must be able to translate performance ratings into action plans for improving or sustaining future performance and must perceive the performance appraisal system as a valuable management tool rather than a curse of supervision. Employees must understand the basis for their evaluations and be motivated to change their behavior in ways more conducive to the attainment of both organizational and individual job outcomes. In short, the validity and utility of every performance rating/appraisal system is dependent on the acceptance of organizational members and their motivation to make the system work as intended.

The present experiment was conducted in conjunction with the development of a behaviorally-based performance appraisal system to replace a set of graphic rating scales based, in part, on traits and unobservable constructs. The old system lacked even face validity (user acceptance) and was pronounced "worthless" by employees and senior management alike. The dearth of field research either comparing behaviorally-based formats or focusing on user

acceptance left the investigators without a sound basis for recommending any given format. Two behavioral rating formats were chosen for evaluation, Smith & Kendall's (1963) behavioral expectation scale (BES) and the behavioral observation scale (BOS) advocated by Latham and Wexley (1981). Company managers and employees evaluated the two behavioral formats in situ for a period of six months, after which a final decision regarding the rating format to support the new appraisal system was to be made.

Since Smith and Kendall's (1963) description of the development and use of BES over 20 years ago, numerous studies have been conducted comparing BES with other rating formats. Such comparitive evaluations have focused almost exclusively on psychometric properties such as interrater reliability, leniency, halo, and convergent and discriminant validity. Reviews of these studies have concluded that there are no differences favoring BES over other methods of performance evaluation (Jacobs, et al., 1980; Kingstrom & Bass, 1981; Landy & Farr, 1980; Schwab, Heneman, & DeCotiis, 1975). However, none of the studies reviewed compared user acceptance of BES versus other behaviorally-based rating formats.

Ronan and Latham (1974) introduced BOS in a study documenting the relationships between the behavior of independent logging crew supervisors and the subsequent performance (e.g., productivity, absenteeism, attendance) of the crews. In similar fashion, BOS have been used in a number of published studies but, in most cases, the method was primarily used to facilitate motivation via goal setting and to document subsequent changes in employee behavior/performance (see Latham & Wexley, 1981). Latham and Wexley (1977) briefly discussed the logic, methodology, and hypothesized benefits of BOS performance ratings. While there have been a few somewhat impassioned exchanges in the literature focusing on BOS vis a vis performance

measurement and feedback (e.g., Bernardin & Kane, 1980; Kane & Bernardin, 1982; Latham, Saari, & Faye, 1980), the most complete explication of BOS as the basis of a performance evaluation system was presented by Latham & Wexley (1981). In contrast to BES, BOS are relatively new and have not been extensively compared with other rating formats. For example, Landy and Farr's (1980) excellent review of the performance rating literature barely mentioned the existence of BOS, devoting only two sentences to the Latham and Wexley (1977) citation. Of the relatively few studies reporting the use of BOS, none have compared user acceptance of BOS versus other behaviorally-based formats.

BES provide an opportunity for raters to record observed behavior relative to scaled anchors, followed by a summary rating on each performance dimension. BOS require the individual rating of specific job relevant behaviors (based on a job analysis) in terms of the relative frequency with which the behavior has been observed by the rater. Performance on each dimension is summarized as the sum (or mean) of the (typically 5 - 15) item ratings. While both formats are based on and document observed job behavior, BOS provide specific evaluative feedback on <u>each</u> behavior (and subsequently, each performance dimension) while BES convey evaluative feedback to the ratee by performance dimension only. Thus, it was hypothesized that the greater behavioral (evaluative) specificity of BOS would be perceived by both supervisors and employees as more accurate and fair, as having greater utility for coaching and development, and as being more motivating than performance ratings/review sessions conducted with the BES format.

Method

<u>Rating Formats</u>

A job analysis for the position of Sales Associate (SA) in a midwestern

retail merchandising chain was conducted via the critical incident technique (Flanagan, 1954). Seventy-three behaviors were identified and placed into nine dimensions by a group of Division Sales Managers (DSMs) using the procedure outlined for the development of BOS dimensions/scales by Latham and Wexley (1981). The behaviors were then formatted as 5-point BOS scales.

Another group of DSMs then scaled the behaviors (Smith & Kendall, 1963); 55 were retained with standard deviations $\leq .5$ and were used to anchor 5-point BES scales. Thus, with the exception of the 18 deleted anchors, the behavioral content of the rating scales was substantially (75%) the same.

Rater Training

The DSMs participated in a mandatory half-day workshop to become familiar with the objectives of the study and the development and implementation of both rating formats. Approximately four hours were devoted to rater training which emphasized the importance of observing and recording employee behavior (Latham, Wexley, & Pursell, 1975; Bernardin & Walter, 1977) and behavior modeling training (Latham, Faye, & Saari, 1979) on conducting a performance review and setting goals (behavioral and numeric) for the next appraisal period. The training program repeatedly emphasized the recording of behaviors (critical incidents) during the entire rating period. DSMs were instructed in recording of their observations on the BES using the anchors to scale their observations before making their ratings (Bernardin & Smith, 1981). On the BOS rating forms, DSMs were told to simply record the relative observed frequency (almost always - almost never) of each behavior listed on the rating forms. DSMs were "walked through" written instructions during training and were provided the same written instructions as part of each packet of performance rating forms.

Design

The 10 stores in the company were approximately matched in terms of the number of DSMs and SAs; one store in each matched pair was randomly assigned to begin the experiment using the BES rating format and the other store the BOS format. All regularly scheduled performance appraisals during the first 3-month period were conducted using the assigned format. During the next three months, each store switched to the other experimental rating format. Thus, order of format usage was counterbalanced across stores and was analyzed in a series of 2 x 2 (format x order) multivariate analyses of variance (MANOVAs).

Subjects and Procedures. A total of 240 non-union SAs were rated on either the BOS or BES appraisal form by their respective DSMs on their regularly scheduled review date. Following the performance rating, the DSMs conducted a formal performance review with each SA, focusing on reinforcement for engaging in effective job behaviors and setting specific behavioral and total point goals for the next scheduled performance review six months later. Following the review the DSMs explained that the company had asked a team of university researchers to develop and evaluate a new performance appraisal system and gave the SAs a questionnaire (5-point response scales) concerning the experimental rating form and the performance review session itself. The SAs were given a preaddressed, postage-paid envelope and were asked to mail the questionnaire directly to the researchers at the university after completing it at their convenience. SAs were assured of the confidentiality of their responses and were asked to write their employee identification number (but not their name) on the form to facilitate comparison of their responses with their performance ratings. Due, in part, to a series of logistics problems (i.e., questionnaires not delivered to stores or to the

SAs), only 165 (69%) of the completed questionnaires were received and included in the analyses.

Results

The 165 completed SA questionnaires were designed to assess several conceptually distinct dimensions of the performance rating/review process. These dimensions were empirically verified by 6 orthogonally-rotated principal components factors with the exception that items concerning perceived fairness and perceived utility for coaching and development all loaded on one factor. The 6 factors, accounting for 75.4% of the variance, are shown in Table 1 with their respective proportions of variance and internal consistency estimates of dimension reliability (coefficient alpha, or Pearson correlation for Factor 3).

Insert Table 1 about here

No significant differences were found in employees' self-ratings of motivation and goal acceptance, the degree of their participation in goal setting, the difficulty of their performance goals, or the degree to which they felt pressured to comply with the demands/expectations of their DSMs (see Table 1). A significant multivariate order effect for goal specificity (3-point scale) indicated that specific goals were perceived as being set more frequently during the first 3 months of the experiment than during the last 3 months, $M = 2.92$ and 2.76, respectively, $F(1, 123) = 3.89$, $p < .05$. This indicates some degree of unevenness in DSMs' implementation of goal setting across time. Because of the counterbalanced experimental design and the absence of an interaction effect, however, the evaluation of the rating formats and the performance review sessions is not biased by this effect.

The major differences between BOS and BES formats occured on Factor 2, Perceived Fairness and Utility for Coaching and Development. The MANOVA

main effect for format was significant, $F(6, 116) = 2.82$, $p < .01$, as were three of the 6 items loading on this factor. All six of the items produced more favorable responses for the BOS format.

Discussion

The data support the hypothesis that the BOS format would be more favorably evaluated by employees than the BES rating format. Employees perceived their performance ratings as more fair, as more accurately reflecting their true level of performance, and they better understood the reasons behind their performance ratings with the BOS rating format. They were also more satisfied with their performance ratings when conducted using the BOS format. Self-rated motivation, effort, and goal acceptance were consistently higher with BOS versus the BES rating format ($p < .03$ by a sign test). No differences between format conditions in the extent to which SAs perceived that specific goals were set, the perceived difficulty of their goals, or their perceived degree of participation in the goal setting process.

An important limitation of these findings is that the observed differences are not the result of direct employee comparisons of BOS versus BES rating formats. The graphic rating scales which had previously been used in this company were the only basis of comparison for either behaviorally-based rating format. The means reported in Table 1 indicate a high degree of acceptance of both behaviorally-based formats compared to the previously used graphic scales. As such, the results comparing BOS versus BES formats in terms of employee acceptance are conservative. That is, without being asked for a direct compairson, employees consistently rated fairness, utility, motivation, and pressure to perform well as greater with the BOS relative to the BES appraisal format (16 out of 16 items, $p < .001$ by a sign test).

We believe that the marked consistency in employee acceptance and

preference for the BOS over the BES rating format is due to the greater degree of behavioral specificity in terms of both feedback and goal setting with the BOS format. The goal setting literature (Locke, 1968; Locke, Shaw, Saari, & Latham, 1981) has convincingly demonstrated that specific goals, if accepted, lead to higher subsequent performance than do general or nonspecific goals. We believe that user acceptance is paramount to making a performance rating/review system work as it should. The greater behavioral (evaluative) specificity of the BOS format combined with the evidence regarding employee acceptance in this study suggests that BOS are likely to be preferred over BES by organizations with respect to the dimensions of performance appraisal and employee motivation systems investigated in this experiment. Of course, these results need to be replicated and extended before sweeping conclusions regarding external validity can be made. We suggest that validity data in terms of subsequent employee performance also should be assessed in future research comparing the BOS and BES rating formats.

References

Bernardin, H. J., & Kane, J. S. (1980). A second look at behavioral observation scales. Personnel Psychology, 33, 809-814.

Bernardin, H. J., & Smith, P. C. (1981). A clarification of some issues regarding the development and use of behaviorally anchored rating scales. (BARS). Journal of Applied Psychology, 66, 458-463.

Bernardin, H. J., & Walter, C. S. (1977). Effects of rater training and diary-keeping on psychometric error in ratings. Journal of Applied Psychology, 67, 64-69.

Cook, D. T., & Campbell, D. T. (1979). Quasi-experimentation: design and analysis issues for field settings. Chicago: Rand McNally.

Flanagan, J. C. (1954). The critical incident technique. Psychological Bulletin, 51, 327-358.

Jacobs, R., Kafry, D., & Zedeck, S. (1980). Expectations of behaviorally anchored rating scales. Personnel Psychology, 33, 595-640.

Kane, J. S., & Bernardin, H. J. (1982). Behavioral observation scales and the evaluation of performance appraisal effectiveness. Personnel Psychology, 35, 635-641.

Kingstrom, P. O., & Bass, A. R. (1981). A critical analysis of studies comparing behaviorally anchored rating scales (BARS) and other formats. Personnel Psychology, 34, 263-290.

Landy, F. J., Barnes, J. L., & Murphy, K. R. (1978). Correlates of perceived fairness and accuracy of performance evaluation. Journal of Applied Psychology, 63, 751-753.

Landy, F. J., & Farr, J. L. (1980). Performance rating. Psychological Bulletin, 87, 72-107.

Latham, G. P., Fay, C. H., & Saari, L. M. (1979). The development of behavioral observation scales for appraising the performance of foremen. Personnel Psychology, 32, 299-311.

Latham, G. P., Saari, L. M., & Fay, C. H. (1980). BOS, BES, and baloney: Raising Kane with Bernardin. Personnel Psychology, 33, 815-821.

Latham, G. P., & Wexley, K. N. (1977). Behavioral observation scales for performance appraisal purposes. Personnel Psychology, 30, 255-268.

Latham, G. P., & Wexley, K. N. (1981). Increasing productivity through performance appraisal. Reading, Massachusetts: Addison-Wesley.

Latham, G. P., Wexley, K. N., & Purcell, E. D. (1975). Training managers to minimize rating errors in the observation of behaviors. Journal of Applied Psychology, 60, 550-555.

Lawler, E. E. (1967). The multi-trait multi-rater approach to measuring managerial job performance. Journal of Applied Psychology, 51, 360-381.

Locke, E. A. (1968). Toward a theory of task motivation and incentives. Organizational Behavior and Human Peformance, 3, 157-189.

Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. Goal setting and task performance: 1969-1980. Psychological Bulletin, 90, 125-152.

Ronan, W. W., & Latham, G. P. (1974). The reliability and validity of the critical incident technique: A closer look. Studies in Personnel Psychology, 6, 53-64.

Schwab, D. P., Heneman, H. G., & DeCotiis, T. A. (1975). Behaviorally anchored rating scales: A review of the literature. Personnel Psychology, 28, 549-562.

Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 47, 149-155.

Table 1

Dimensions and Mean Format Ratings on

Sales Associate's Performance Appraisal Questionnaire

	Mean Rating BES	Mean Rating BOS	F	df
Factor 1 - Employee Motivation and Goal Acceptance (30.8%, _ = .93)	4.49	4.61	.31	5, 115
1. How personally committed are you to achieving your performance rating goal at your next performance rating?	4.47	4.60	.81	1, 119
2. How much effort do you plan to expend in attempting to achieve your performance goal?	4.54	4.65	.61	
3. To what extent have you accepted your goal for the next performance rating?	4.34	4.51	1.46	
4. How hard do you plan to work in order to achieve your goal on the next performance rating?	4.41	4.55	.97	
5. To what extent do you really want to (i.e., for personal satisfaction) achieve your performance rating goal next time?	4.68	4.76	1.00	
Factor 2 - Fairness and Utility for Coaching and Development (16.0%, _ = .89)	4.03	4.35	2.82***	6, 116
1. To what extent do you think the performance rating reflected your true performance?	3.81	4.31	9.66***	1, 121
2. To what extent do you think the performance rating was fair?	3.90	4.36	8.07***	
3. To what extent did the performance rating help you and your DSM identify areas for coaching which will improve your performance?	4.11	4.33	1.71	
4. To what extent do you understand the reasons behind the performance rating you received?	4.11	4.45	3.28*	
5. How satisfied are you with your performance rating?	3.93	4.29	4.87**	
6. How satisfied are you with the performance review session?	4.29	4.33	.11	

Table 1 (Cont.)

	Mean Rating BES	Mean Rating BOS	F	df
Factor 3 - Participation in Goal Setting (10.0%, r = .41) (High scores reflect greater personal influence)	3.11	3.28	1.35	2, 118
1. Who set the performance rating goal for your next performance rating?	2.84	2.81	.00	1, 119
How much influence did you have in setting your next performance rating goal?	3.38	3.74	2.42	
Factor 4 - Goal Difficulty (7.0%, _ = .56)	2.50	2.66	.56	3, 118
1. How difficult do you think it will be for you to achieve your goal on the next performance rating?	2.28	2.47	.98	1, 120
2. How challenging do you think your performance rating goal is?	3.72	3.95	1.32	
3. How confident are you that you can achieve your goal at the next performance rating?	1.51	1.56	.23	
Factor 5 - External Pressure (6.3%)				
1. To what extent do you plan to work toward your goal only because your DSM wants you to work toward it?	3.13	3.22	.06	1, 121
Factor 6 - Goal Specificity (5.3%) (3-point scale, 3 = yes)				
1. Did you and/or your DSM set a specific goal for you to "shoot for" on your next performance rating?	2.86	2.85	.28	1, 123

Note. The df for all univariate F-ratios within a factor are identical and are shown only for the first item in a factor.

*p < .10. **p < .05. ***p < .01.